Game data pipeline:
Building vs buying

When considering data analytics for your games, should you build your own data pipeline or venture into buying a proven solution?

Considering building your own data pipeline?

Building your own data pipeline can be appealing to studios and publishers seeking full control and customization. However, it comes with significant challenges and considerations that often outweigh the benefits.

Initially, building your own pipeline requires substantial upfront investment in infrastructure, technology, and personnel. This includes hiring skilled engineers and data scientists, as well as investing in hardware, software licenses, and ongoing maintenance costs. These expenses can quickly escalate, especially when factoring in the complexities of data storage, processing, and compliance with evolving privacy regulations.

Developing a custom pipeline demands time and expertise to design, implement, and optimize. It involves navigating technical hurdles such as data integration, scalability issues, and ensuring reliability under varying workloads. This process not only consumes resources but also diverts focus from core game development and strategic initiatives.

Moreover, maintaining an in-house pipeline involves ongoing operational overhead. This includes monitoring system performance, troubleshooting issues, and continuously updating infrastructure to meet evolving data demands. These responsibilities can strain internal resources and detract from innovation and competitive differentiation in the fast-paced gaming market.

While the idea of control and customization may be tempting, the practical benefits and cost efficiencies of purchasing a data pipeline often outweigh the risks and expenses associated with building and maintaining one internally - especially for studios and publishers who are yet to reach their peaks.

The benefits of buying a third-party data solution

Opting to purchase a data pipeline, rather than building one in-house, provides major strategic advantages for game developers, studios and publishers. Firstly, **a third-party solution offers standardized event schemas and predefined KPI calculations** across all games in your portfolio. This consistency ensures that metrics are comparable and decision-making is data-driven across projects.

Moreover, vendor-managed solutions benefit from economies of scale in data storage and processing. This translates into **cost efficiencies for your business**, with pricing typically structured around Monthly Active Users (MAU), **simplifying cost forecasting** compared to variable cloud costs.

A third-party data pipeline also reduces time to insight, with the set-up taking a matter of hours or days instead of months. This **reduces the workload on Data Analysts and Scientists**, allowing them to focus on strategic analysis rather than routine data preparation.

Unlike relying on an (often small) team of in-house experts, third-party solutions provide **scalability and continuity**. They eliminate the risk of knowledge silos and dependencies on individual engineers, ensuring operational stability and long-term sustainability of data operations.

Finally, purchasing a data pipeline **shifts the majority of the responsibility of data privacy and compliance to the vendor**. Third-party data pipelines also include **robust data governance frameworks** that regulate access rights.

This ensures authorized personnel have timely access to the right data, enhancing **security and compliance** with data privacy regulations.

> "With GameAnalytics we have a highly modular C# SDK that handles A/B tests, cross-promotion events, in-app purchases, push notifications, and so on. PipelineIQ is really flexible and scalable over time, which means we can keep incorporating new events into our dashboards."
>
> TapNation   Read full story

Set up requirements

When considering options, purchasing a pre-built solution like GameAnalytics' PipelineIQ Pro offers a quicker setup with immediate access to robust analytics tools and integrations, minimizing initial setup time and resource investment while providing long-term efficiency and scalability.

On the other hand, building a data pipeline allows for additional customization and control over architecture and data handling processes. However, this option requires a longer period of time to set up and investment to design, implement, and maintain the infrastructure.

Here is an estimated outline of the set up requirements for both options:

Building your own pipeline

Initially, you'll need to make architectural decisions, determining the framework and technologies that will support your pipeline. This often involves selecting a cloud infrastructure provider like AWS, Google Cloud, or Azure, based on your specific needs and preferences.

Once the cloud infrastructure is set up, the next step is creating event schemas. This involves defining the structure and format of the data points you want to collect from your game, ensuring consistency and clarity in the information gathered. Concurrently, defining event longevity rules is essential; this determines how long different types of data will be retained for analysis and compliance purposes.

Instrumenting events into the game itself is a critical phase where Game Engineers integrate code to capture relevant data points during gameplay. This step requires careful implementation to ensure accurate and comprehensive data collection, without impacting game performance or user experience.

Following event instrumentation, data processing becomes pivotal. This includes setting up pipelines and workflows to ingest, transform, and store the collected data in a way that facilitates efficient querying and analysis. Quality assurance (QA) processes are integral throughout this stage to validate data accuracy, completeness, and integrity.

Documenting ETL processes and defining key performance indicators come next. Documentation ensures transparency and continuity in data operations, while KPIs provide measurable benchmarks for evaluating game performance and player behavior.

Finally, data visualization and interpretation complete the setup, enabling stakeholders to derive actionable insights from the collected data. This step involves using BI tools or custom dashboards to visualize trends, patterns, and correlations within the data, facilitating informed decision-making and optimization of game strategies.

Building your own data pipeline requires above all expertise in cloud computing, data engineering, and game development - making it a complex but customizable solution tailored to specific organizational needs and data governance requirements.

> "GameAnalytics SDK doesn't cause any conflicts at all. It's so seamless and light. I've never seen any errors or crashes, so I'm really happy with this tool."
>
> Ivan Panasenko,
> CEO at Creauctopus
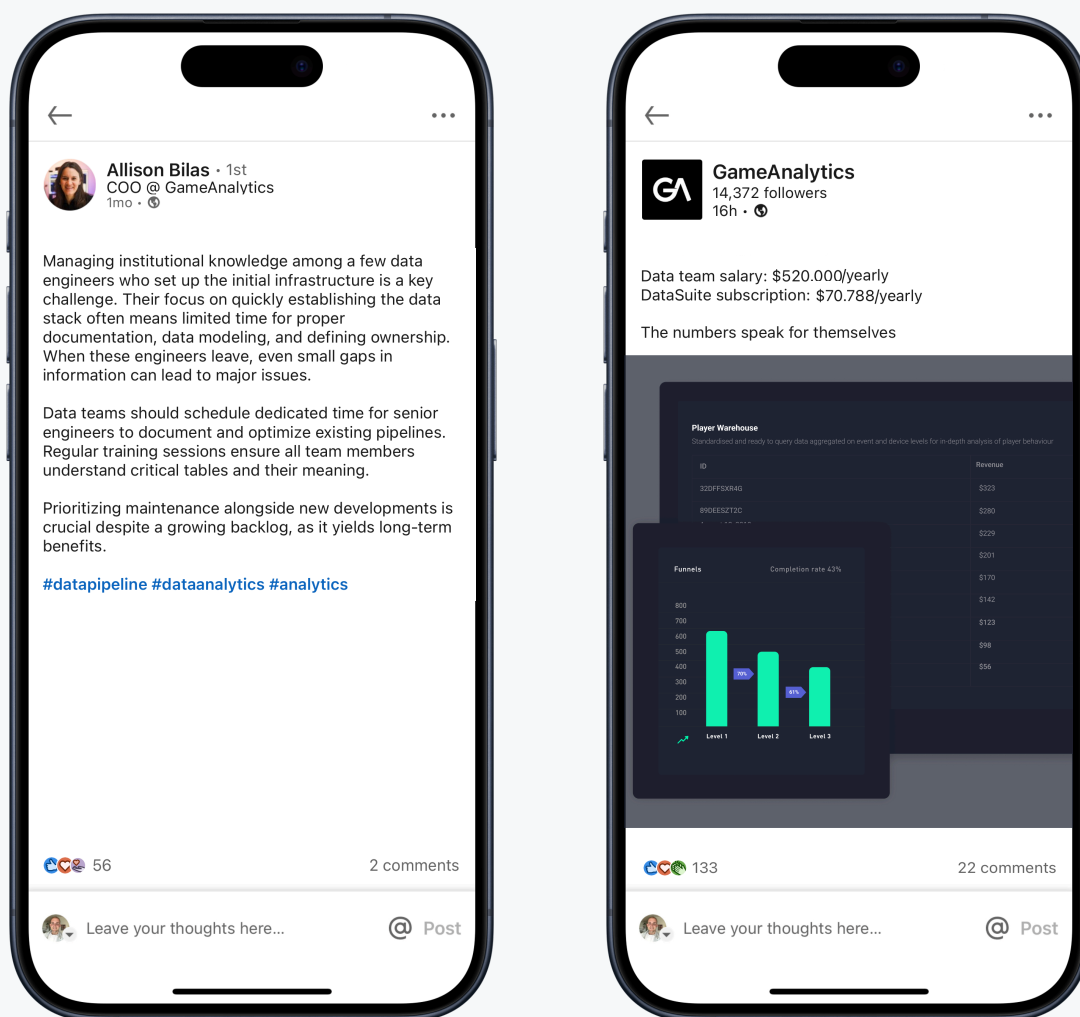
Getting started with PipelineIQ Pro

There are two solutions provided as part of GameAnalytics' PipelineIQ Pro: Data Warehouse and Data Export.

To set up Data Warehouse, we require a Google Group email address from you. We'll handle the rest and set up the BigQuery project for your data. After providing the email, there might be a short wait period before the setup is fully activated.

For Data Export, we'll need details from your cloud storage solution (like AWS or Google Cloud) such as the bucket ARN/ name, region, and configured permissions. Once you've specified the type of data format you want to send (JSON or Parquet), you should start seeing data flowing in 15-30 minutes for JSON and by the next morning for Parquet.

For both solutions, an SDK integration is necessary to connect data from your game(s) to GameAnalytics. With one line of code, you can begin capturing relevant events and player interactions - ready to be queried.

DataSuite is designed to democratize data access and analysis, making it accessible and actionable for users across various roles within your organization. You don't need specialized data skills to use PipelineIQ effectively. Basic SQL knowledge is advantageous for querying data, and tools like Google Data Studio empower you to create insightful reports and dashboards without coding.

Example scenario and pricing

To showcase how your costs could be calculated, the items to look out for and what your estimated **Total Cost of Ownership (TCO)** could be, we prepared an example scenario.
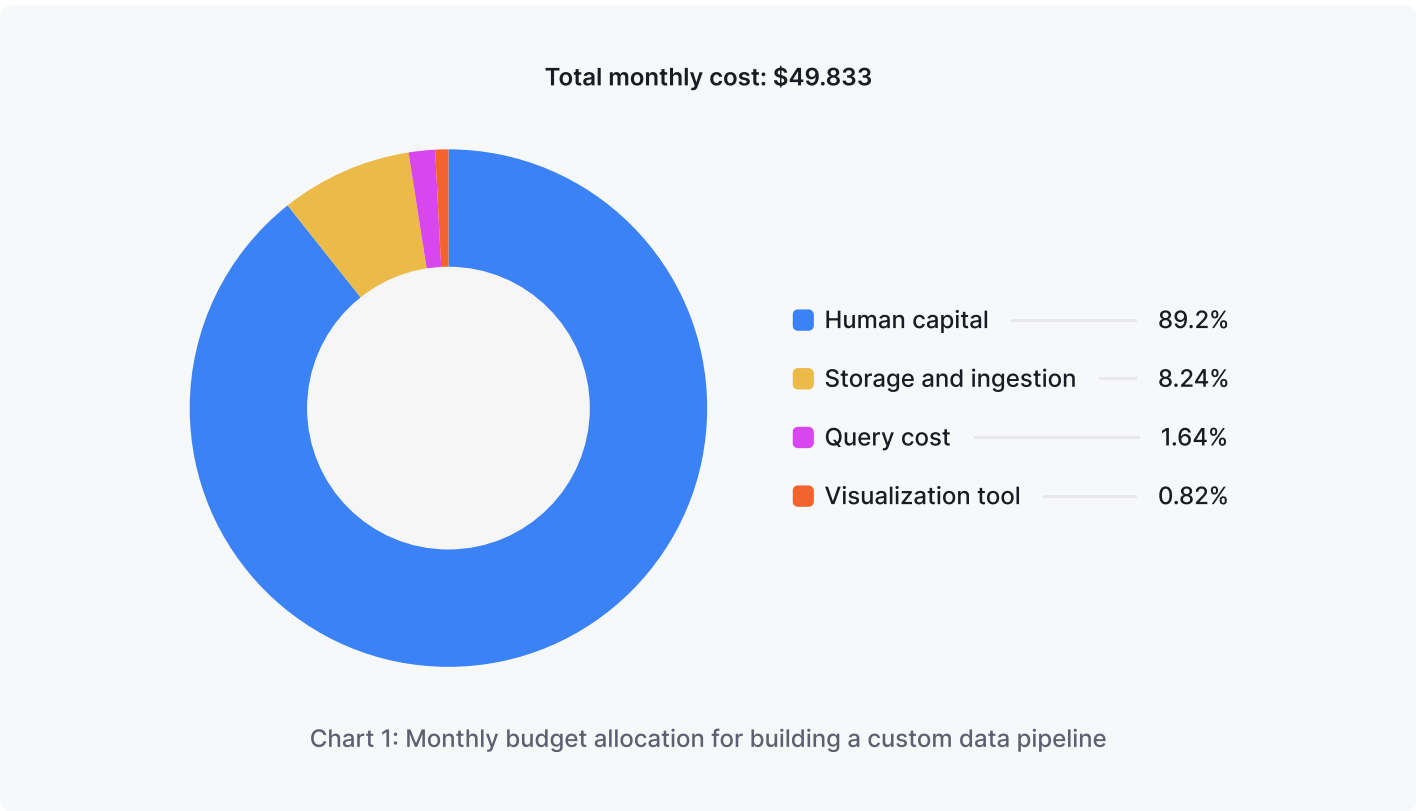
Let's say you run a mid-sized game studio that recently launched it's first game. Alongside you, the Studio Director, your small team consists of a few Developers and Engineers, Game Designers, a Product Manager, Growth Manager, and Customer Support. Luckily, your game is becoming a hit and rapidly gains 5 million monthly active users (MAU). Initially focused on game development, content updates, and basic marketing strategies, your studio's success prompts a realization: the need for a data-driven approach to further propel your games success. You now have two options:

1. Build

Currently, your funds are secure and you do not have to make data-driven decisions immediately. You are ready to hire a new data team focused on building custom data infrastructure, ongoing maintenance and conducting analyses. This team will consist of a Data Engineer, a Data Scientist, an Analytics Team Lead and a Data Analyst. Your budget allows for an average yearly salary of $130.000 per role.

Your monthly ingestion and storage cost per KMAU is $1, totaling to $5.000 as your current MAU is 5.000.000 (keep in mind that this number will scale as your audience grows over time). Additionally, you need to account for the query costs. While the price may differ, we estimate at least $1.000 monthly.*

Visualization tools are usually priced per seat. Based on the number or viewers, the cost can easily reach $500 or more.

Total monthly cost: $49.833

🔵 Human capital	89.2%
🟡 Storage and ingestion	8.24%
🟣 Query cost	1.64%
🟠 Visualization tool	0.82%

Chart 1: Monthly budget allocation for building a custom data pipeline

This includes the cost of essential components for building a data pipeline such as data ingestion, processing, orchestration, storage, warehousing, and analytics/visualization platforms. Infrastructure costs for maintaining the data pipeline include expenses for data transfer, processing / ETL / query costs, storage, and visualization software. Finally, other costs can arise from issues like unreliable data pipelines, spaghetti SQL, orphaned tables, technical debt, security and privacy compliance, and time spent.
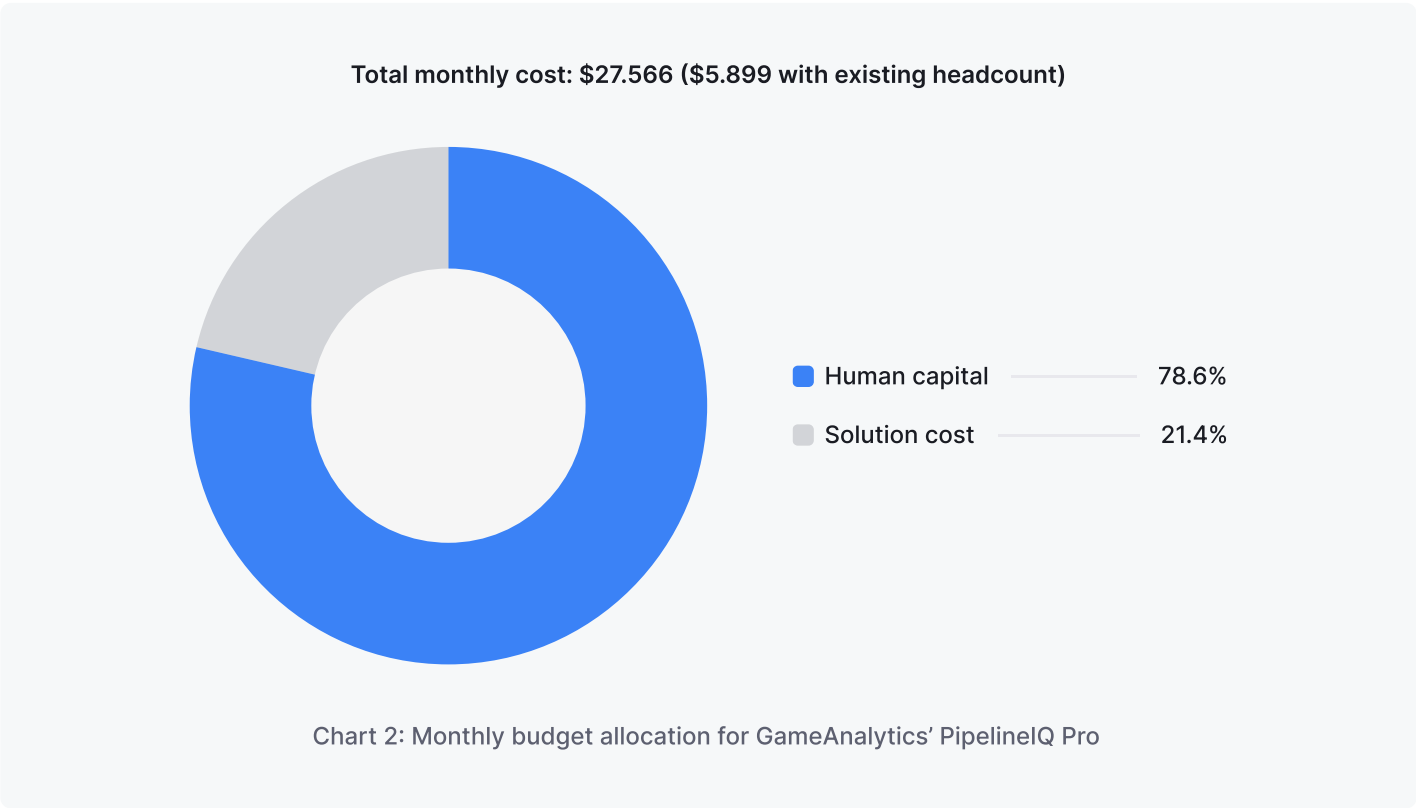
2. Buy

You want to make sure that your success hypothesis is correct and wish to back your decisions with data.

You realize that you do not have the budget, time or personnel to build an in-house solution, so you begin to look for a third-party solution. After coming across GameAnalytics, you realized that PipelineIQ Pro fits your needs for a data pipeline that can be up and running in no time.

Currently, your budget allows you to expand your team with a Data Analyst and Data Scientist, with a yearly salary of $130.000 each. Although you have these or similar roles on your team already, you are aware that allocating new data-related responsibilities to them could be challenging. As your game keeps growing, you are looking to implement analytics as soon as possible.

Buying GameAnalytics' PipelineIQ costs you the fee of a monthly subscription of $5.899, with transparent scaling pricing based on MAU and an out of the box infrastructure ready for you.

Total monthly cost: $27.566 ($5.899 with existing headcount)

● Human capital	78.6%
● Solution cost	21.4%

Chart 2: Monthly budget allocation for GameAnalytics' PipelineIQ Pro

Expense conclusion

In both scenarios, the primary costs revolve around human capital. Employing the right people with the necessary expertise and skills. However, opting for vendor-managed solutions can significantly reduce these costs.

GameAnalytics' PipelineIQ Pro is fully developed and maintained by us, eliminating the need for ongoing efforts. Moreover, existing team members often fit well into managing these tools, making integration and operations more efficient without the need for additional recruitment.

Buying a third-party solution not only lowers initial setup costs but also minimizes long-term resource requirements, making it a pragmatic choice for studios looking to optimize their data infrastructure.

What should your team look like when building your own data pipeline?

Legend:
- Data Analyst (magenta)
- Data Scientist (blue)
- Data Engineer (yellow)
- Game Engineer (orange)

Radar chart axes: Data ETL, Database management, Data tools, Data visualization, Storytelling, Business insights, Metrics & Reporting, Experimentation, Inference, Stats & ML modeling, Collection mechanism, Events schemas

Role	Description
Data/Analytics Team Lead	This role ensures the efficient and effective flow of data within the organization. They oversee the design, development, and maintenance of the data pipeline infrastructure, collaborating closely with cross-functional teams to understand data requirements and ensure alignment with business objectives.
Data Engineer or BI Developer	They are building and managing the data pipeline to ensure the timely, accurate, and complete flow of data within the organization. They are responsible for instrumenting and collecting events from various sources, transforming raw data into meaningful insights that drive business decisions.
Data Analyst	Specializing in SQL or other relevant languages, their role revolves around working with data to provide insights into what is happening within each game. Their expertise lies in extracting, transforming, and analyzing data to generate reports that connect metrics to key business outcomes. Typically, an analyst can manage 1 to 2 games, with workload allocation dependent on factors such as the stage of the games' lifecycle and the complexity of gameplay and analysis requirements.
Data Scientist	This role is pivotal in leveraging advanced analytics techniques to provide predictive and prescriptive insights aimed at forecasting likely future outcomes and understanding the key drivers that impact these outcomes within each game. With a focus on data-driven decision-making, data scientists use statistical modeling, machine learning algorithms, and data mining techniques to extract actionable insights from complex datasets.
Game Engineer	Their primary responsibility is to instrument events within the game to trigger data collection, a critical task that requires meticulous attention to detail and extensive collaboration with the development team. This involves implementing tracking mechanisms, defining event schemas, and establishing communication channels with backend systems for data transmission.

What to consider when you're looking for a data pipeline for your games

In any infrastructure decision, understanding the Total Cost of Ownership (TCO) is key. TCO encompasses not only the direct costs of technology and infrastructure, but also the investment in human resources required to build, manage, and optimize that infrastructure. This tends to make up the majority of the TCO, as it includes the salaries for skilled engineers, data scientists, and support staff. This holistic view of the cost underscores the importance of considering long-term operational expenses, alongside initial setup costs, when evaluating whether to build or buy a data pipeline.
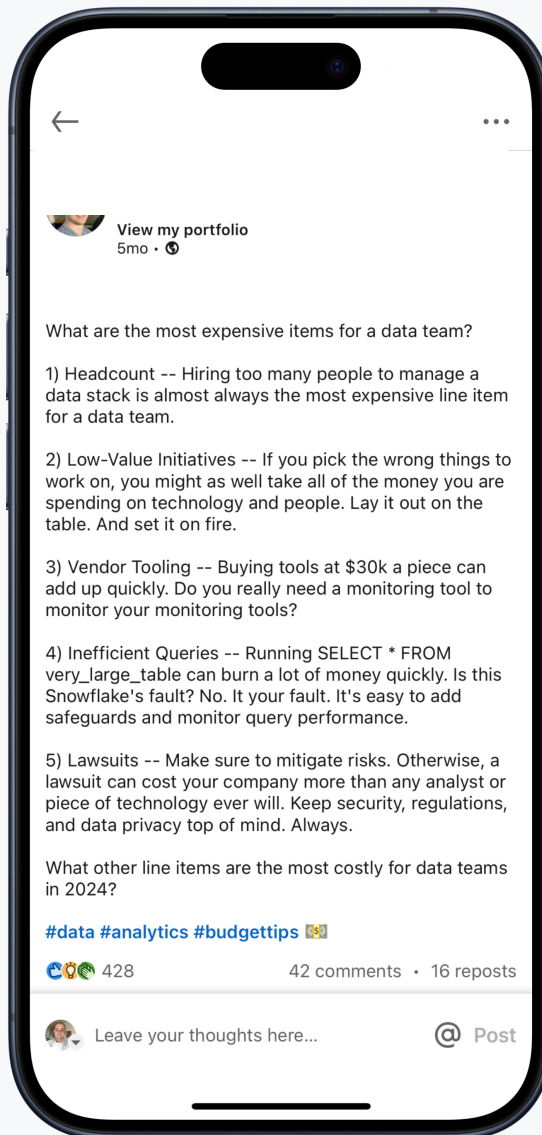
	Build	Buy
In-house roles	When you're looking to build a data pipeline in-house, these are the roles you're likely to need: Data/Analytics Team Lead Data Engineer or BI Developer Game Engineer Data Analyst Data Scientist	A third-party solution streamlines operations with pre-configured settings, eliminating the need for extensive hiring. Two roles that you might need include: Data Scientist Data Analyst
Expenses	The price of building a solution is dependent on the amount of data you're looking to process and the level of analysis you expect. You also must account for the future scaling of your needs and additional costs connected to your growth. Your estimated price: $49.833 monthly	Opting for a third-party solution strips your costs down to the price of the solution and staffing. Your estimated price: $27.566 monthly *Note: $5.899 with existing headcount*
Set-up requirements	Building a data pipeline for an existing game from scratch can take between 9 to 12 months.	Setting up a third-party solution typically takes days, instead of months.

> "You'd usually have to have another Data Engineer to sort out the analytics pipeline. Using a data warehouse means that you can just hire one person to manage one tool and use that raw data."
>
> Happy Volcano   Read full story

View my portfolio
5mo

What are the most expensive items for a data team?

1) Headcount -- Hiring too many people to manage a data stack is almost always the most expensive line item for a data team.

2) Low-Value Initiatives -- If you pick the wrong things to work on, you might as well take all of the money you are spending on technology and people. Lay it out on the table. And set it on fire.

3) Vendor Tooling -- Buying tools at $30k a piece can add up quickly. Do you really need a monitoring tool to monitor your monitoring tools?

4) Inefficient Queries -- Running SELECT * FROM very_large_table can burn a lot of money quickly. Is this Snowflake's fault? No. It your fault. It's easy to add safeguards and monitor query performance.

5) Lawsuits -- Make sure to mitigate risks. Otherwise, a lawsuit can cost your company more than any analyst or piece of technology ever will. Keep security, regulations, and data privacy top of mind. Always.

What other line items are the most costly for data teams in 2024?

#data #analytics #budgettips 💵

428   42 comments • 16 reposts

Leave your thoughts here…   @ Post

PipelineIQ Pro from GameAnalytics

PipelineIQ Pro is a versatile package of tools that help you access your raw data, player-level data, and aggregated metrics from the GameAnalytics platform, freeing you from the constraints of black-box analytics solutions.

Use our suite of data tools to power-up your insights with custom data analysis, including churn prediction, LTV calculations, and user-level segmentations.

Data Warehouse

A fully managed data workbench. Data Warehouse brings all your game analytics into one BigQuery instance, pre-processed and ready for advanced analysis. Run SQL queries, drill into event and player data, and compare insights across multiple sources.

Data Export

Raw game data delivered straight to you. Data Export streams all events in JSON or Parquet directly to your AWS or Cloud Storage bucket in real time, giving you full control to process and integrate the data however you want.

Metrics API

Metrics API gives you programmatic access to KPIs like DAU, retention, and ARPU, so you can automate reporting and track performance across all your games. Paired with the Organization API, you can also keep studios, games, and users updated automatically without touching the AnalyticsIQ UI.

Game Data Sharing

Securely access partner data through Data Warehouse, Data Export or Metrics API. Game Data Sharing makes it easy to query and compare KPIs across studios, enabling deeper collaboration and growth opportunities.

Ready to get started? <u>Get in touch</u> and book a demo.

GameAnalytics

GameAnalytics.com
LinkedIn
Facebook